

CODE EXCITED LINEAR PREDICTIVE SPEECH ENCODING DEVICE

Patent Number: JP8123493
Publication date: 1996-05-17
Inventor(s): HOSAKA SACHIKO; KATAOKA AKITOSHI; MORIYA TAKEHIRO; HAYASHI SHINJI
Applicant(s):: NIPPON TELEGR & TELEPH CORP <NTT>
Requested Patent: ☐ JP8123493
Application Number: JP19940264235 19941027
Priority Number(s):
IPC Classification: G10L9/14 ; G10L9/18 ; H03M7/30
EC Classification:
Equivalents:

Abstract

PURPOSE: To provide the code excited linear predictive speech encoding device which obtains not only a speech, but also a natural decoded sound even if a background noise, music, etc., are added to the speech.
CONSTITUTION: The code excited linear predictive speech encoding device is equipped with a shape excitation source code book, which consists of plural subordinate code books; and the subordinate code books 106a and 106b consist of speech parts 106a-1 and 106b-1 consisting of signal source vectors learnt by using speech containing no noise and non-peech parts 106a-2 and 106b-2 consisting of signal source vectors learnt by utilizing non-speech signals other than the speech and random signal source vectors that are not learnt. Consequently, the quality is improved as compared with when the subordinate code books are composed of only the vectors of the speech parts 106a-1 and 106b-1 or non-speech parts 106a-2 and 106b-2.

Data supplied from the **esp@cenet** database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-123493

(43) 公開日 平成8年(1996)5月17日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 9/14	G			
9/18	E			
H 0 3 M 7/30	Z	9382-5K		

審査請求 未請求 請求項の数 2 O L (全 7 頁)

(21) 出願番号 特願平6-264235

(22) 出願日 平成6年(1994)10月27日

(71) 出願人 000004226

日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号

(72) 発明者 保坂 様子

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内

(72) 発明者 片岡 章俊

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内

(72) 発明者 守谷 健弘

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内

(74) 代理人 弁理士 志賀 正武

最終頁に続く

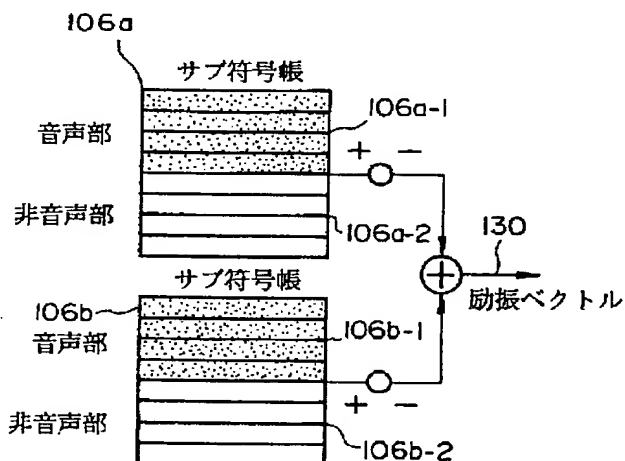
(54) 【発明の名称】 符号励振線形予測音声符号化装置

(57) 【要約】

【目的】 音声のみならず、音声に背景雑音や音楽等が加わっても自然な復号音を得られる符号励振線形予測音声符号化装置を提供すること。

【構成】 形状励振源符号帳を備える符号励振線形予測音声符号化装置において、形状励振源符号帳は、複数のサブ符号帳から構成され、各サブ符号帳(106a, 106b)は、雑音を含まない音声を用いて学習した信号源ベクトルから構成される音声部(106a-1, 106b-1)と、音声以外の非音声を利用して学習した信号源ベクトル又は学習を行わないランダム信号源ベクトルから構成される非音声部(106a-2, 106b-2)とからなることを特徴とする。

【効果】 サブ符号帳を音声部のみあるいは非音声部のみのベクトルで構成するよりも品質を向上させることができる。



【特許請求の範囲】

【請求項1】 形状励振源符号帳を備える符号励振線形予測音声符号化装置において、

前記形状励振源符号帳は、複数のサブ符号帳から構成され、

前記各サブ符号帳は、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を用いて学習した信号源ベクトルから構成される音声部と、音声以外の非音声を利用して学習した信号源ベクトル又は学習を行わないランダム信号源ベクトルから構成される非音声部とからなることを特徴とする符号励振線形予測音声符号化装置。

【請求項2】 形状励振源符号帳を備える符号励振線形予測音声符号化装置において、

前記形状励振源符号帳は、第1および第2のサブ符号帳から構成され、

前記第1のサブ符号帳は、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を用いて学習した信号源ベクトルのみから構成され、前記第2のサブ符号帳は、音声以外の非音声を利用して学習した信号源ベクトルのみ又は学習を行わないランダム信号源ベクトルのみで構成されることを特徴とする符号励振線形予測音声符号化装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は、 $2\text{ kbit/s} - 16\text{ kbit/s}$ 程度のビットレートで使われる符号励振線形予測(CELP)音声符号化方式、残差駆動線形予測(RELP)音声符号化方式などの、音声合成フィルタを励振信号源で駆動する形式の音声符号化方式に適用し、音声のみならず、音声に背景雑音や音楽等が加わっても自然な復号音が得られる符号励振線形予測音声符号化装置に関するものである。

【0002】

【従来の技術】近年、ディジタル移動通信などの技術分野においては、電波を有効利用するなどの目的で、種々の高能率符号化方式が用いられている。 8 kbit/s 程度の符号化速度で音声を符号化する高能率符号化方式の一つに、符号励振線形予測(CELP)符号化方式がある。この符号励振線形予測符号化方式は、人間の音声の生成機構をモデル化する線形予測分析合成にベクトル量子化技術とA b S (合成を利用した分析)技術を組み合わせた周知の技術である。また、符号励振線形予測符号化方式のなかには、2つのサブ符号帳によって形状励振源符号帳を構成した、共役構造CELP (CS-CELP)と呼ばれる構造のものがある(なお、共役構造CELPについては、本願出願人の出願による特願平-70534「音声の符号化方法」に詳細に説明されている)。

【0003】ここで、図3を参照して、共役構造を持つ

符号励振線形予測符号化方式の一例について説明する。図3は、共役構造CELP符号化方式による符号励振線形予測音声符号化装置の一構成例を示すブロック図であり、この図において、符号1は、入力端子であり、この入力端子1からは、アナログの音声信号をサンプリング周波数 8 kHz でサンプリングして生成したデジタルの音声データが入力される。入力端子1から入力された入力信号は、LPC(線形予測)分析器2へ入力される。このLPC分析器2は、入力信号を、窓かけ、自己相関関数計算、斉次連立方程式の求解等の周知の手法によって、線形予測分析し、音声合成フィルタ4の予測係数と音声分析フィルタ3の予測係数を求める。そして、音声合成フィルタ4の予測係数(線形予測係数)は、一度量子化され、伝送に適した形に変換された後、再度復号されて、音声合成フィルタ4に設定される。また、音声分析フィルタ3の予測係数も、同様に設定される。

【0004】音声分析フィルタ3は、音声合成フィルタ4における合合作用と逆の作用をなすフィルタであり、音声合成フィルタ4の出力信号(複合音声)を、LPC分析器2によって設定された予測係数に基づいて分析することによって、線形予測残差信号を求める。そして、線形予測残差信号に基いて、音声の基本周期(ピッチ)成分をあらわす適応励振源符号帳5の記憶データが設定される。この適応励振源符号帳5は、ピッチ周期時間とピッチパルスの振幅をピッチ周期ベクトルとして表し、記憶した符号帳であり、音声のピッチ変化に追従して適応的に変化するものである。

【0005】6aおよび6bは、形状励振源符号帳を構成するサブ符号帳であり、残差信号からピッチ成分を除いた残りの波形である形状励振源成分を、励振ベクトルとして表し、記憶した符号帳である。サブ符号帳6a、6bからなる形状励振源符号帳は、非周期的励振に対応するものであり、時間と共に変化しない。

【0006】そして、適応励振源符号帳5ならびにサブ符号帳6a、6bからは、後述する最小歪み計算部8によって選択されたピッチ周期ベクトルおよび励振ベクトルが取り出され、適応励振源符号帳5から取り出されたピッチ周期ベクトルには、利得部10において、ピッチゲインが乗算され、一方、サブ符号帳6a、6bから取り出された各励振ベクトルは、互いに加算されて励振ベクトル30として合成された後、利得部11において、ゲインアダプタ9によって設定された予測ゲインが乗算され、さらに利得部13において、形状ゲインが乗算される。

【0007】そして、利得部10の出力ベクトルと利得部13の出力ベクトルは、互いに加算された後、音声合成フィルタ4に供給され、音声合成フィルタ4において、上述したようにしてLPC分析器2によって設定された線形予測係数に基づいて合成される。なお、上記の利得部10におけるピッチゲインは、ピッチ周期の励振

に対する振幅のゲインであり、音声合成フィルタ4に入力されるベクトルに応じて設定されるものである。他方、利得部11における予測ゲインは、形状励振に対する振幅のゲインであり、ゲインアダプタ9において、過去の励振ベクトル30aのパワーに基づいて線形予測分析を行うことによって設定されるものである。

【0008】そして、入力端子1から入力された入力信号ベクトルから、音声合成フィルタ4の出力ベクトルが減算されて歪データが求められ、この歪データが、聴覚重み付けフィルタ7において、人間の聴覚の特性に対応した係数によって重み付けされた後、最小歪み計算部8へ入力される。そして、最小歪み計算部8において、聴覚重み付けフィルタ7から出力された歪データのパワーが計算され、この歪データのパワーが最も小さくなるように適応励振源符号帳5ならびにサブ符号帳6a、6bから、それぞれピッチ周期ベクトルおよび励振ベクトルが選択される。なお、通常、上記の聴覚重み付けフィルタ7は、移動平均自己回帰型の10次程度のフィルタであり、フォルマントの山の部分をやや強調するような特性を持つように構成されたものであり、最小歪み計算部8は、2乗誤差最小の計算を行うように構成されたものである。

【0009】そして、符号出力部12において、上述した予測係数、ピッチ周期ベクトルおよび励振ベクトルそれぞれに対して選択されたコード、利得等が、ビット系列の符号に変換され、さらに必要に応じて訂正符号が付加され、符号出力部12から伝送路へ向けて出力される。すなわち、符号化の際は、入力信号の波形に対し、合成波形（音声合成フィルタ4の出力ベクトル）の聴覚重み付け自乗誤差が最小となるような励振源の組み合わせが、適応励振源符号帳5ならびにサブ符号帳6a、6bからなる形状励振源符号帳から選ばれるのである。

【0010】上述したように、共役構造CELP（CS-CELP）において、形状励振源符号帳は、2つのサブ符号帳6a、6bに分かれて構成されている。各サブ符号帳6a、6bは、相互に他を補完する機能を果たし、仮に伝送路の符号誤りにより、復号器（図示せず）において、一方のサブ符号帳の形状励振ベクトルが誤って選ばれても、他方のサブ符号帳のベクトルとの和をとるため、結果として大きな誤差を生じないという特徴がある。

【0011】ところで、サブ符号帳6a、6bからなる形状励振源符号帳に記憶される形状励振源成分を求める際にも、上述した適応励振源符号帳5を設定する場合と同様に、ベクトル量子手法が適用される。しかし、音声の変化に応じて形状励振源符号帳を適応的に変化させるには、あまりにも分散が大きいため、形状励振源成分を求める場合には、予め尤もらしい波形をベクトルとして複数用意して、符号帳に記憶させておくことが行われる。ただし、伝送の際、実際に伝送されるのは、ベクトル

ルの指標のみである。

【0012】この尤もらしい符号帳を構成する手法の一つが学習である。図3に示す形状励振源符号帳の場合、雑音を表すベクトルの一群を初期状態として、予め音声データベースを用いて学習がなされる。学習は、各励振源に合成フィルタを畳み込み、振幅（ゲイン）を乗じて得た合成波形と目標音声信号との誤差の、符号帳中の全ての励振源についての総和が最小となるよう各形状励振源符号帳のサブ符号帳のベクトルを決定するものである。

【0013】ここで、図4を参照して、従来の形状励振源符号帳の学習においてサブ符号帳に設定される励振ベクトルの構成について説明する。通常の音声データベースは、背景雑音や背景音楽等を含まないため、このような音声データベースを用いた場合、学習済みの形状励振源符号帳は、図4（a）に示すように、それを構成する2つのサブ符号帳6a-1、6b-1が、共に雑音を含まない音声データの信号源ベクトルからなる音声部のみで構成される。一方、学習をしない場合（例えばガウス雑音からなる符号帳の場合）、あるいは、音声以外すなわち非音声である背景音声等を利用して学習した符号帳の場合、図4（b）に示すように、2つのサブ符号帳6a-2、6b-2が、共に学習を行わないランダム信号源ベクトルもしくは音声以外の非音声を利用して学習した信号源ベクトルからなる非音声部のみで構成される。

【0014】

【発明が解決しようとする課題】背景雑音や背景音楽等を含まない音声データベースによって学習済みの形状励振ベクトル信号すなわちサブ符号帳6a-1、6b-1の出力から合成した励振ベクトル30-1（図4（a）参照）は、雑音の無い入力音声に最適なものとなる。したがって、結果的に、このような学習済みの符号帳を有する符号化装置は、雑音の無い入力に対して高品質であるが、音声の背景に雑音や音楽などが加わると極めて明らかな品質劣化を生じるという問題点があった。

【0015】一方、学習をしない、例えばガウス雑音からなる符号帳あるいは音声以外の非音声である背景音声等を利用して学習した符号帳から合成される励振ベクトル30-2（図4（b）参照）は、雑音のある場合の入力音声に最適なものとなる。したがって、付加雑音による品質劣化が少ないものの、雑音の無い環境下の音声に対してあまり品質が上がらないという欠点があった。

【0016】本発明は、上記の問題点に鑑みてなされたものであり、音声のみならず、音声に背景雑音や音楽等が加わっても自然な復号音が得られる符号励振線形予測音声符号化装置を提供することを目的とする。

【0017】

【課題を解決するための手段】請求項1記載の発明は、形状励振源符号帳を備える符号励振線形予測音声符号化

装置において、前記形状励振源符号帳は、複数のサブ符号帳から構成され、前記各サブ符号帳は、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を用いて学習した信号源ベクトルから構成される音声部と、音声以外の非音声を利用して学習した信号源ベクトル又は学習を行わないランダム信号源ベクトルから構成される非音声部とからなることを特徴とする。

【0018】また、請求項2記載の発明は、形状励振源符号帳を備える符号励振線形予測音声符号化装置において、前記形状励振源符号帳は、第1および第2のサブ符号帳から構成され、前記第1のサブ符号帳は、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を用いて学習した信号源ベクトルのみから構成され、前記第2のサブ符号帳は、音声以外の非音声を利用して学習した信号源ベクトルのみ又は学習を行わないランダム信号源ベクトルのみで構成されることを特徴とする。

【0019】

【作用】以上の構成によれば、各サブ符号帳を構成する音声部と非音声部のベクトルの分布が異なっているため、音声部と非音声部をうまくカバーするように両者から選んでサブ符号帳を構成することにより、雑音を含まない音声入力に対しても背景音楽を含む音声入力に対してもそれぞれ最適なベクトルが選択されるため、総合的に考えるとサブ符号帳を音声部のみあるいは非音声部のみのベクトルで構成するよりも品質が向上する。

【0020】

【実施例】以下、図面を参照してこの発明による一実施例を説明する。なお、本発明は、図3を参照して説明した符号励振線形予測符号化方式等に設けられている形状励振源符号帳の構成に係り、その初期値の設定に特徴があるものである。したがって、符号励振線形予測音声符号化装置のそれ以外の構成については、従来のものと同様であり、以下の説明では、形状励振源符号帳の構成について詳細に記述する。

【0021】図1は、この発明による符号励振線形予測音声符号化装置の形状励振源符号帳の構成を示す構成図であり、一例として、図3に示すものと同様に、2つのサブ符号帳を有する場合の形状励振源符号帳を示したものである。この図に示すサブ符号帳106a、106bは、図3に示すサブ符号帳6a、6bに対応するものであり、それぞれを構成するベクトルは、実際の構成において、例えば、各40次元で128個であるが、この図においては、それぞれを8個のベクトルによって簡略して表している。

【0022】サブ符号帳106a、106bを学習させるのに先立ち、まず、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を使って学習することによって得られた符号帳と音声以外の非音声例えば付加雑音や背景音楽を使って学習することによって得られ

た符号帳を用意する。

【0023】次に音声入力（例えば車の中）を考慮し、（音声部のベクトル個数）+（非音声部のベクトル個数）=128個になるように予め用意した符号帳の音声部と非音声部のベクトルの中から適切なベクトルを選ぶことによって、各サブ符号帳106a、106bが、それぞれ音声部106a-1または106b-1と非音声部106a-2または106b-2を持つ複合形状の符号帳となるように、サブ符号帳106a、106bそれぞれに音声部のベクトルと非音声部のベクトルを記憶させる。

【0024】各128ベクトルの内、例えば64ベクトルを音声部、残り64ベクトルを非音声部に振り分けることができる。ただし、この比率は1対1に限らず、使用状況に応じた入力音声の性質を見込んで自由に設定することができる。

【0025】符号励振線形予測音声符号化装置の実際の動作においては、上述したように、各サブ符号帳106a、106bの中のベクトルの選択に当たって、聴覚重みづけを考慮して歪最小となるベクトルが選ばれるから、一方、雑音を含まない入力音声の場合は、2つとも音声部（106a-1および106b-1）のベクトルが選ばれ、励振ベクトル130が合成され、他方、入力音声为背景音声等を含んでいる場合には2つとも非音声部（106a-2および106b-2）が選ばれて励振ベクトル130が合成される可能性が高い。したがって、従来の場合すなわち背景雑音や背景音楽等を含まない音声データベースによって学習済みサブ符号帳を用いて音声の背景に雑音や音楽などが加わった入力信号を符号化する場合、あるいは、学習をしない符号帳もしくは音声以外の非音声である背景音声等を利用して学習した符号帳を用いて付加雑音による品質劣化が少ないものもしくは雑音の無い環境下の音声を符号化する場合と比較して、励振ベクトル130（図3に示す励振ベクトル30に対応するもの）として、入力信号の特性により適合したものが選択されることになるので、音声品質を向上させることができる。

【0026】次に、本発明による他の実施例を図2を参照して説明する。この図において、形状励振源符号帳は、図1に示す各サブ符号帳106a、106bに代えて、音声部のみからなるサブ符号帳206aと非音声部のみからなるサブ符号帳206bから構成されている。この場合、サブ符号帳206bを構成する非音声部は、雑音波形の時系列のベクトルまたは雑音や非音声信号からなるデータベースを用いたベクトル128個から構成されたものであり、一方、サブ符号帳206aを構成する音声部は、同様の雑音ベクトルを初期値として、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方のデータベースを用いて学習した結果得られた128個の励振ベクトルから構成されたものである。

【0027】この図に示すようにして各サブ符号帳206a、206bを構成した場合、各サブ符号帳が、それぞれ128個の音声部あるいは非音声部からなる励振ベクトルによって構成されるので、図1に示した場合と比較して、一方、サブ符号帳206aでは、付加雑音による品質劣化が少ないものもしくは雑音の無い環境下の音声に対して適切な励振ベクトルの選択の範囲が広がり、他方、サブ符号帳206bでは、音声の背景に雑音や音楽などが加わった入力信号に対する適切な励振ベクトルの選択の範囲が広がる。したがって、各サブ符号帳から選択された励振ベクトルの合成ベクトルである励振ベクトル230（図3に示す励振ベクトル30に対応するもの）は、図1に示す実施例と同様、従来の場合に比較し、結果として音声品質の向上を図ることができる。

【0028】なお、以上の実施例では、形状励振源符号帳を2つに分けたCS-CELPの例を示したが、更に多数の符号帳に分ける構成にも、また、1つのみの符号帳を用いる場合も本複合符号帳の構成手法を適用することができる。

【0029】また、符号帳のメモリを削減するため、一つまたは数個のベクトルを循環的に使用方法も矛盾なく組み合わせることができる。また、演算量削減のため符号帳ベクトルの非重要成分を0として、数点の代表パルスのみを残すスパース励振ベクトルの採用も本方式と併用することが出来、効果を損なわない。このとき、音声部符号帳のみにスパース励振ベクトルを適用し、非音声部符号帳はすべての標本に値を持つように設定することが可能で、品質を保ちつつ演算量やメモリを削減出来る点で実用的に有効である。

【0030】さらに、雑音励振部分の検索において数個の励振パルスを順次予め決った位置に立てて歪みの少ないパルス列を求める代数的符号励振CELP（A-CELP）と組み合わせても効果を発揮することは明らかである。

【0031】本発明を適用した場合を、従来技術による場合（形状励振源符号帳に全て音声部ベクトルを用いた場合、及び、非音声部ベクトルのみを用いた場合）と比較した効果を以下に示す。

○従来技術

（音声部ベクトルのみの場合）全ての励振音源を学習した方式では、雑音を含まない入力音声に対して好ましい品質を示すものの、背景に雑音の混じった音声では、背景の雑音が変形し違和感のある妨害音となって聴感上の劣化がはげしい。

（非音声部ベクトルのみの場合）入力信号が雑音のない音声、雑音や音楽などが含まれる音声に対しても重大な差のない品質が得られる。ただし、同一のビットレートで学習音源を用いたものを、本発明による一実施例と比較した場合、復号音声の信号対雑音比で1.0-2.0dB程度劣り、8kbit/s程度の符号化では、聴感

上ざらざらした雑音感がぬぐえないという実験結果が得られた。

【0032】○本発明

背景に雑音のないときは最適な音声部のベクトルが選択されるため、音声部のみと同程度の品質感が得られ雑音感は避けられる。また、背景雑音が存在するときは最適な非音声部ベクトルが選択されるため、雑音が変化して違和感をあたえることを避けられる。よって、広範囲の入力音声条件に対して高品質な音声の符号化が可能である。

【0033】

【発明の効果】請求項1記載の発明によれば、入力信号が雑音を含まない音声であれば従来構成例と同様な音声部のベクトルが選ばれるので、従来に比べ劣化がない。また、入力信号に付加雑音や背景音楽などを含んでいれば、非音声部符号帳のベクトルが選ばれるので、従来に比べ品質を向上させることができる。すなわち、音声部と非音声部のベクトルの分布が異なっているため、音声部と非音声部をうまくカバーするように両者から選んでサブ符号帳を構成することによって、雑音を含まない音声入力に対しても背景音楽等を含む音声入力に対してもそれぞれ最適なベクトルが選択されるため、総合的に考えるとサブ符号帳を音声部のみあるいは非音声部のみのベクトルで構成するよりも品質が向上する、という効果を得ることができる。

【0034】また、請求項2記載の発明によれば、雑音を含まない音声入力に対しては、雑音を含まない音声又は雑音を含まない音声と雑音を含む音声の両方を用いて学習した信号源ベクトルのみから構成された第1のサブ符号帳から最適なベクトルが選択され、背景音楽等を含む音声入力に対しては、音声以外の非音声を利用して学習した信号源ベクトルのみ又は学習を行わないランダム信号源ベクトルのみで構成された第2のサブ符号帳から最適なベクトルが選択されるので、総合的に考えると2つのサブ符号帳を音声部のみあるいは非音声部のみのベクトルで構成するよりも品質を向上させることができる。

【図面の簡単な説明】

【図1】本発明の一実施例による形状励振源符号帳の構成を示す構成図である。

【図2】本発明の他の実施例による形状励振源符号帳の構成を示す構成図である。

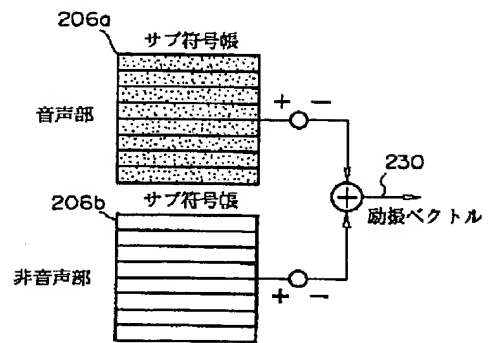
【図3】従来の符号励振線形予測音声符号化装置の構成を示すブロック図である。

【図4】従来の形状励振源符号帳の構成を示す構成図である。

【符号の説明】

106a、106b、206a、206b サブ符号帳
106a-1、106b-1 音声部ベクトル
106a-2、106b-2 非音声部ベクトル

【図 2】



(b)

フロントページの続き

(72)発明者 林 伸二

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内